

Yinkun Qu

(312) 532-4785 | yinkunq@gmail.com | [LinkedIn](#) | [Github](#)

EDUCATION

University of Illinois, Chicago

Expected: December 2026

Master of Science, Electrical and Computer Engineering

GPA: 3.6/4.0

- Relevant Coursework: Computer Vision, Neural Network, Computer Architecture, Parallel Processing.

Xi'an University of Architecture and Technology, China

September 2021 - July 2025

Bachelor of Computer Science and Technology

GPA: 3.7/5.0

- Coursework: Data Structures & Algorithms, Web Development, Operating Systems, Software Engineering, OOP.
- Awards: Second Prize at the National College Students Mathematical Modeling Contest, 2023; Second Prize at the National College Students Computer Design Contest, 2023; Scholarships: Striving for Excellence, Second Prize, 2021–2023.

SKILLS

Certifications: Oracle Cloud Infrastructure 2025 Certified AI Foundations Associate

Languages: Java, Python, C, C#, Javascript, MPI

Frameworks: Spring Boot, Spring MVC, MyBatis, jQuery, .NET, FastAPI, Distributed Systems, Node.js

Data & Caching: MySQL, Redis, PostgreSQL, SQL query optimization, SQL lite

DevOps & Cloud: Docker, Kubernetes, Linux, Nginx, Git/GitHub, Maven, Gradle, JUnit, Shell scripting, AWS, Azure, CI/CD

AI/Agent: LangChain, LLM APIs, RAG, multi agent workflows, context management, prompt engineering, LLM evaluation, PyTorch, LangGraph, Function calling, Embeddings, MCP

INTERNSHIP

Jinhailong Technology Co., Ltd. China

February 2025 - April 2025

Software Implementation & Operations Engineer at Cross-border Trade / Customs System

- Built 8+ stored procedures, 6 triggers, 12+ views for order/finance dashboards; replaced manual spreadsheets, pre-shipment checks reduced from 15 to 5 min/order which is **67% faster than before**.
- Integrated Kingdee **ERP** with strong reservation; wrote SQL stored procs/triggers for pre-lock & auto-reconciliation, late shipments **25% less than before**.
- Helped operations: basic monitoring and backup drills, role-based field permissions, runbooks, and mobile approvals.

PROJECT

Local First AI Vault Assistant

February 2026 - present

- Designed and developed a local-first **RAG** knowledge assistant for Markdown notes, supporting document parsing, text chunking, vector indexing, semantic retrieval, context-augmented question answering, and citation-grounded responses to improve personal knowledge retrieval and reuse.
- Built a LangChain-based RAG pipeline with Markdown loading, chunking, embedding generation, vector retrieval, prompt assembly, and grounded answer generation; enriched each chunk with metadata such as file path, heading hierarchy, and last modified time to improve retrieval interpretability and traceability.
- Designed a controllable **Agent workflow** using LangGraph, decomposing question analysis, knowledge retrieval, context selection, answer generation, citation validation, and self-checking into separate workflow nodes to improve maintainability and debuggability of complex LLM applications.
- Implemented **tool-calling** capabilities including search_notes, read_note, extract_todos, and suggest_backlinks, enabling the Agent to search notes, read source content, extract action items, and recommend bidirectional links for knowledge management workflows.
- Built backend services and deployment infrastructure with FastAPI, Redis, PostgreSQL, and Docker, supporting asynchronous indexing tasks, query caching, index status management, and basic RAG evaluation, with extensible design for cost monitoring, retrieval optimization, and multi-user access control.

Mobile Health & Sports Application

March 2025 - June 2025

- Designed and implemented a front-end/back-end separated health management system supporting both a web admin portal and mobile clients.
- Developed core backend features (user auth with JWT, role-based access, and CRUD for recipes/workouts) and optimized APIs to achieve less than **200ms average response time**.
- Delivered video/image upload and streaming module supporting files **up to 1000MB**, improving content richness.
- Completed the full software engineering cycle (requirements, design, implementation, testing) in **4 months**.

Open-Source Contribution

CERCA - [Citation Extraction & Reference Checking Assistant](#)

Java MVC, JUnit, Java FX, Maven

- Added unit tests for CERCA's PDF reference extraction pipeline using sample PDFs, improving coverage of core parsing functionality, designed edge-case tests improving exception handling and module robustness.
- Contribute to reference verification function make sure average reference extracting and verrificating time < 500ms